International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

MACHINE LEARNING PROJECT ON EARTHQUAKE PREDICTION SYSTEM

Amit Kumar
School of Computing Science and Engineering
Galgotias University, India
Kumaramit95543094 @gmail.com

Manas Gupta
School of Computing Science and Engineering
Galgotias University, India
manasgupta1807@gmail.com

ABSTRACT– In current situation, an explicit forecaster is planned and created, a framework that may conjecture the disaster. It centres around distinguishing early indications of Earthquake by utilizing AI calculations. Framework is qualified for essential strides of making learning frameworks alongside life pattern of data science. Informational indexes for Indian sub-mainland alongside remainder of the globe are gathered from government sources. Pre-handling of data is trailed by development of stacking model that joins Random Forest and Support Vector Machine Algorithms. This mathematicalmodel is formed using algorithms hooked in to “training data- set”. Model searches for design that prompts fiasco and go with it in its structure, to create decisions and gauges without being explicitly modified to play out the assignment. After figure, we broadcast the message to government authorities and across different stages. Time, Locality and Magnitude, these 3 factors addresses the focus of information acutely.

Keyword: SVM, Random Forest, training data-set.

I. INTRODUCTION

Earthquakes are natural accidents that can occur on land or underwater and can cause tidal or land earthquakes causing a lot of damage. Depending on the size, this can also lead to death. Earthquakes are widely observed and are implicated in many academic studies, so only basic concepts are described here. Most seismic activity occur s between the movements of lithospheric plates (tectonic plates). This action consists of accumulating energy in the pressing pose of the stone, so that it is released suddenly.
After an earthquake, the environment (longitude, magnit ude and depth), time and magnitude are determined.
Magnitude is the physical size of an earthquake, so the energy released can also be roughly estimated by conver ting the instantaneous magnitude. Via seismic events and via auxiliary effects such as avalanches, crevices, avalanches, flames and waves. Earthquakes cause destru ction and loss of life. Having the ability to stage these rare events will help reduce the damage done, with actions such as warning residents and government offes.

II. LITERATURE REVIEW

As far as I know, there are 2 studies that attempt to use m achine learning to predict when the next earthquake will occur. Both extrapolate that predicting the next event is extremely difficult due to randomness and the difficulty of proving that earthquakes follow specific patterns. Imp ortantly, both surveys used recorded seismic vibration tables to generate machine learning models. Other ML applications were also investigated: a study specifically focused on the prediction of aftershock events, which follow larger events and is a crucial topic because after shocks can still cause many injuries, yielded good results There is a discussion of the information science methodo logy used. Laboratory earthquake experiments were studi ed with ML facilities, which attempted to predict failure times.Another article examines the design of energy sign als, fom lowabundance seismic waves to the planning of slow-sliding events.

III. TOOLS/TECHNOLGY USED

- Google Collab
- Pandas – 1.2.4
- matplotlib.pyplot
- Neural Networks
-

Some libraries are needed to import first import

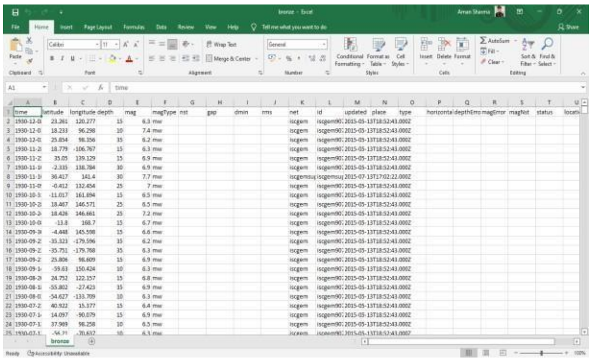
numpy as npy

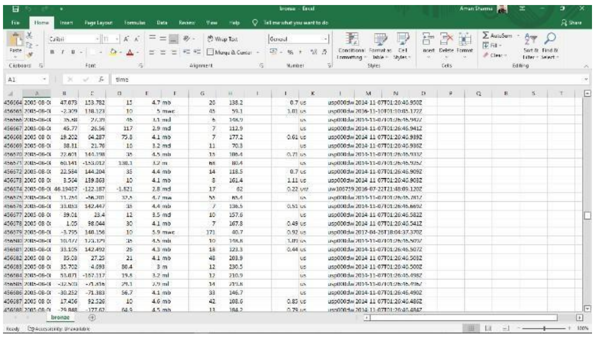
import pandas

import matplotlib.pyplot as pltl

Dataset

Datasets we used are obtained from two websites USGS.gov and IMD.gov. Datasets have 6 Columns: Date, Latitude, Longitude, depth, time, Longitude. till Till the end of 2018 all the records of whole world were downloaded from these websites and later we filtered these record as portrayed below:





IV. PROPOSED SYSTEM

Creating predicting modelling includes slow methodology. Tools and technology which are expectedly utilized for creating model are Python.

Support Vector Machine Vs Random Forest

A Support Vector Machine is a Supervised Machine Learning Which can be used for both Classification and Regression task . In Classification, SVM is used to classify data points into different classes .The goal is to find best separate data points into high-Dimensional space

```
[ ] from sklearn import metrics
    print("model performance accuracy: ", metrics.accuracy_score(y_test, y_predictions))

    model performance accuracy:  0.9733333333333334
```

Random Forest Algorithm

Random Forest Algorithm is widely used in machine learning that can be applied both classification anad regression Random Forest is Known for its ability to handle complex datasets and produce result .It is Important to use original language and provide proper references when discussing technical concepts to maintain acaemic integrity.

```
from sklearn import metrics
print("Accuracy:",metrics.accuracy_score(y_test, y_predictions))

Accuracy: 0.9666666666666667
```

A. Data Acquisition

The process of collecting data from relevant sources for production use. Data can be obtained from sources external and internal to the system, or from data produced by the system. Potential advancements may begin and involve gathering the required information. The datasets we use come from two websites USGS.gov and IMD. government. The dataset has 6 columns: date, latitude, longitude, depth, time, longitude.

B. Data Pre-Processing

Data preprocessing is the way to prepare information and fit it to machine learning models. This is the critical and main step in building a model using machine learning.

Data Engineering:
Real world data is not an actionable and organized structure, it may turn out to be false, invalid, false, misleading, information may be lost. Unnecessary and dubious information can logically hinder design approval and disclosure of information during the preparation phase. So this is the biggest development in ML systems and a need for data cleaning to remove these features or approve/fix them. This includes data integration, handling of partially missing values, error correction, care and transformation of categorical values.

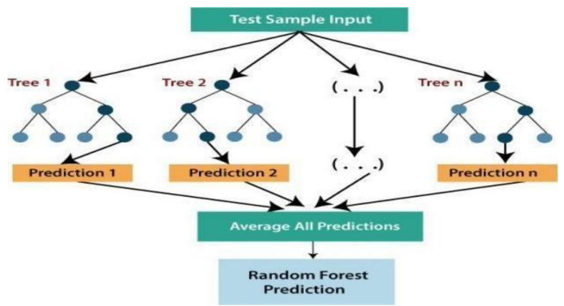
Feature Engineering:
A method to select and modify the most important factor from raw data using domain knowledge when developin g predictive models via ML. The main objective is to improve the performance of ML algorithms. There are 4 stages in feature engineering: feature creation, feature transformation, feature extraction, and feature selection. The dataset contains many features that are not very use ful for prediction. So basically feature engineering reduces those features and provides a set of useful features that help in inaccurate/accurate predictions. It reduces storage footprint, execution time and data.

V. MODELLING

First, the target variables and the a variables are underst ood and realized. Second, split the dataset into training a nd test datasets, and third, create a regressor/classifier model and fit it to the training dataset. In python, scikit- learn can be a simple, basic, and efficient open source library that implements a distribution of machine learnin g algorithms including different classification, regressio n, and clustering algorithms with a unified interface.

Building A Random Forest Regression Model

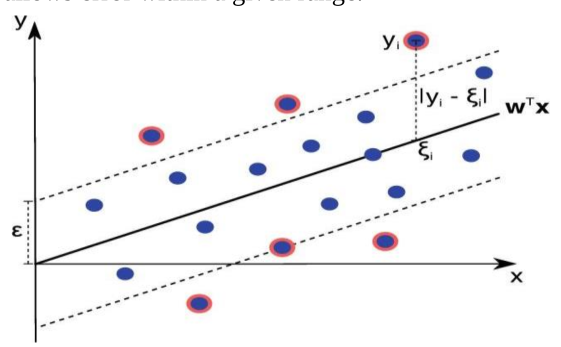
Random Forest is an equipped learning technique that can be created for regression and classification tasks. It un dertakes the task of constructing multiple decision tree s during training and outputs the class, either the avera ge prediction of each individual tree (regression) or the model of the class (classification). These trees form a forest. Decision trees are rule-based models; Given a set of training data with goals and characteristics, a decision tree algorithm will suggest classification and regression rules. The vertices are the nodes of the tree, their presence and absence will solve the probability problem.It is useful to have a regular appr oach to understand it. Base nodes and split nodes are base d on the information gain index. Therefore, a random for est can be a model containing different trees, with the pot ential to create decision support rules, moreover, the tech nique for selecting root and parent nodes israndom.



Building A Support Vector Machine RegressionModel

Regression and classification tasks will be performed by support vector machines (a supervised learning algorithm).

SVMs separate different classes of data using decision lines called hyperplanes. When predicting a numerical value, SVR tries to find the function f(x) in the style of the decision limit at a specific deviation from €, which can be a threshold in all prediction horizons, obtained from the target value Yi, the initial hyperplane, specified by The data points are within the defined range. This decision limit is the tolerance range - the limit that allows error within a given range.



Building A Stacking Regressor Mode

Stacking regression is an ensemble learning method. Several regression models collaborate, as a result, meta-regressor is build & itself finds its best fit by making use of output of individual regression models, trained on absolute training set, as meta- features. Widely wont to attain accuracy.

Result

Stacking model is most accurate with 84% compared to other two Boosting and Bagging. For large datasets, the random forest- SVM (Support Vector Machine) works well. For all the methods Response time is equal. Stacking took much higher training time. See below table for results:

Parameters / Algorithms	ACCURACY PERCENTAGE	TRAINING TIME	RESPONSE TIME
Bagging	73%	3m3sec	5 sec
Boosting	77%	3m21sec	5 sec
Stacking	84%	11m36sec	5 sec

VI. FUTURE SCOPE

Moving standard deviations are the accompanying bunch of highlights to be tried. Adding more data, very much like the flux from IMOs, can maybe work on the outcomes. the data must be obtained, changed and designed to confirm assuming there's any secret example that could work on the outcomes. Different areas is explored, and move learning could be plausible.

VII. CONCLUSION

Accordingly, we can infer that that integration of seismic actions with machine learning technology produces effective and critical outcome and may be wont to foresee earthquake broadly, given the previous history of the indistinguishable is very much kept up with. Our endeavour are frequently named effective. The cooperation of the 2 can additionally be progressed to intensely safeguard quakes more. Huge datasets influence be extremely critical. Prediction models are much of a time deployed in a section centric manner, thus increasing the probabilities of accurate prediction exponentially but at the price of studying algorithms wont to build Stacking model, since it will perform well giving the algorithms decided to make meta regressor are exact themselves. the use of the technique might be extended in foreseeing different nature tragedies moreover.

REFERENCES

[1] Wonliorld Health Organization, Earthquakes, Health topics.

[2] Sameer, Earthquake History (1965–2016): Data Visualization and Model Development (2019), Medium.

[3] DeVries, P.M.R., Viégas, F., Wattenberg, M. et al. Deep learning of aftershock patterns following large earthquakes (2018), Nature 560, 632–634.

[4] Mignan, A., Broccardo, M. One neuron versus deep learning in aftershock prediction (2019), Nature 574, E1–E3.

[5] B. Rouet-Leduc, C. Hulbert, N. Lubbers, K. Barros, C. J. Humphreys, P. A. Johnson, Machine Learning Predicts Laboratory Earthquakes (2017), Geophysical Research Letters 44, 9276–9282

[6] P. A. Johnson, B. Rouet-Leduc, L. J. Pyrak-Nolte, G. C. Beroza, et al., Laboratory earthquake forecasting: A machine learning competition (2021), Proceedings of the National Academy of Sciences, 118.

[7] Synced, Harvard & Google Seismic Paper Hit With Rebuttals: Is Deep Learning Suited to Aftershock Prediction? (2019), Medium.

[8] R. Shah, Stand Up for Best Practices: (2019), Medium.